# STRUCTURE FROM VISUAL MOTION AS A NONLINEAR OBSERVATION PROBLEM [1]


**S. SOATTO* R. FREZZA** P. PERONA* ****

**California Institute of Technology 116-81, Pasadena–CA 91125. E-mail soatto@systems.caltech.edu*

***Università di Padova, Dipartimento di Elettronica. Via Gradenigo 6/A 35100 Padova–Italy.*



**Abstract.** Over the last decade, estimating scene structure from visual motion has become a central task of computational vision. As it is well known, the estimation task is *nonlinear* due to the perspective nature of the measurements. One may ask whether there exists a smart choice of coordinates that simplifies the estimation task. In particular, since "linearity" is a coordinate-dependent notion, one may seek for a particular choice of coordinates such that the problem of estimating structure from motion becomes linear and spectrally assignable. Unfortunately, such a choice of coordinates does not exist, even if we allow for a nonlinear change of output coordinates or an embedding into a higher-dimensional state-space. As a consequence of this result, we study some alternative estimators with nonlinear error dynamics which are proved to converge, and legitimate the use of local linearization-based techniques for estimating structure from known motion and visual information. In most of the cases, however, the true motion undergone by the viewer is unknown. We propose a novel dynamic estimator for scene structure which is independent of the motion of the viewer. The method consists in the identification of an Exterior Differential System with parameters on a sphere.



**Key Words.** Visual structure estimation, structure from motion, identification of exterior differential systems, implicit Extended Kalman Filter.


## 1. INTRODUCTION

Estimating "Structure From Motion" (SFM) consists of reconstructing the structure of a moving object from its projection onto a camera. A number of schemes have been proposed for estimating motion from known structure, structure for known motion, and both structure and motion recursively from an image sequence (see (Faugeras, 1993; Zhang and Faugeras, 1992) for a review of the existing methods). In this paper we restrict our attention to the *recursive* estimation of *point-based* structure from motion; we will first consider the case of *known motion*, and then the case in which the motion is unknown.

It has been known for a while (Matthies *et al.*, 1989) that SFM can be formulated in terms of the estimation of the state of a nonlinear dynamical system. Such estimation task has been traditionally approached by using Extended Kalman Filters (EKF) (Jazwinski, 1970), as for example in (Matthies *et al.*, 1989; Oliensis and Inigo-Thomas, 1992; Shekhar and Chellappa, 1992; Soatto *et al.*, 1993).

### 1.1. *MODELING STRUCTURE FROM MOTION*

Let us simplify the problem by assuming that the motion of the objects being viewed is *rigid, constrained on a plane*, and has *constant velocity*. It can be shown (Soatto, 1994) that the planar motion case is structurally equivalent to the full 3D motion, as far as observability is concerned. The "structure" of the scene is represented by a number of point-features whose coordinates in the ambient plane are $\mathbf{x} \doteq [\mathbf{x}_1\ \mathbf{x}_2]^T$; $v = [v_1\ v_2]^T$ indicates the relative translational velocity between the object and the viewer in the body frame and $\omega$ is the rotational velocity about an axis orthogonal to the ambient plane and to the optical axis. We assume that each point-feature corresponds to a line which is perpendicular to the viewing line (i.e. we see a horizontal slice of a cylindrical world). The horizontal coordinate of the *projection* of the point onto an image plane is $y \doteq \mathbf{x}_1/\mathbf{x}_2$, so that we can write a nonlinear dynamical model having the position of the point in the ambient plane as the state, and the projection as the measured output:

$$\begin{cases} \frac{d}{dt}\mathbf{x} = f(\mathbf{x}) & \mathbf{x}(t_0) = \mathbf{x}_0 \in \mathbb{R}^n \\ y = h(\mathbf{x}) \end{cases} \tag{1}$$

where

$$\begin{cases} n = 2 \\ f(\mathbf{x}) \doteq \hat{\Omega}\mathbf{x} + V \\ \hat{\Omega} \doteq \begin{bmatrix} 0 & -\omega \\ \omega & 0 \end{bmatrix} \\ V \doteq \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ h(\mathbf{x}) \doteq \frac{\mathbf{x}_1}{\mathbf{x}_2} & \mathbf{x}_2 \neq 0. \end{cases} \tag{2}$$

We call the above model the *standard model for SFM*. Estimating structure from motion is then equivalent to estimating the state of the above model. One may argue that the choice of the reference frame (the viewer reference in this case) and of the model of projection (an ideal pinhole camera with unit focal length) are arbitrary. We fully agree. There are other possible reference frames (object-centered, world-centered etc.) and models for the perspective


[1] Research sponsored by NSF NYI Award, NSF ERC in Neuromorphic Systems Engineering at Caltech, ONR grant N00014-93-1-0990.

projection (with the center of projection displaced in the ambient plane). More than that, there are other possible *nonlinear* changes of coordinates, not simply changes of the reference frame, that one may consider. Therefore, it may be interesting to study whether any of these changes of coordinates *structurally simplifies* the estimation problem. An estimator (observer) for the model (1) is defined as a nonlinear dynamical system of the form

$$\frac{d}{dt}\hat{\mathbf{x}} = g(\hat{\mathbf{x}}, y) \qquad \hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0$$

which has the measurements $y$ as inputs and produces the estimates of the state of the original model. The error $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ also satisfies a set of *nonlinear* differential equations. Unlike the linear context, it is not easy in general to design observers such that the estimation error has prescribed dynamical properties. Over the last ten years (see for instance (Hermann and Krener, 1977; Nijmeijer, 1982; Schaft, 1982; Krener and Isidori, 1983; Krener and Respondek, 1985; Levine and Marino, 1986; Lee and Nam, 1991; Phelps, 1991)), various techniques have been developed to study the existence of a coordinate transformation of $\mathbb{R}^n$

$$\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n \tag{3}$$

$$\mathbf{x} \mapsto \mathbf{z} = \Phi(\mathbf{x}) \tag{4}$$

such that the original model is transformed into the so-called "observer form"

$$\begin{cases} \frac{d}{dt}\mathbf{z} \doteq \left[\frac{\partial \Phi}{\partial \mathbf{X}} f(\mathbf{x})\right]_{\mathbf{X}=\Phi^{-1}(\mathbf{Z})} = A\mathbf{z} + \alpha(C\mathbf{z}) \\ y = h(\Phi^{-1}(\mathbf{z})) = C\mathbf{z} \end{cases} \tag{5}$$

for some $A$, $C$ such that the pair $(C, A)$ observable; $\alpha$ is a smooth function of the outputs. If such a $\Phi$ can be found, then an observer of the form

$$\frac{d}{dt}\hat{\mathbf{z}} = (A + LC)\hat{\mathbf{z}} - Ly + \alpha(y) \tag{6}$$

yields an estimation error $\mathbf{e} \doteq \mathbf{z} - \hat{\mathbf{z}}$ satisfying the differential equation

$$\frac{d}{dt}\mathbf{e} = (A + LC)\mathbf{e} \tag{7}$$

that is *linear and spectrally assignable* through an appropriate choice of the gain matrix $L$.

## 2. UNFEASIBILITY OF THE THE OBSERVER LINEARIZATION

We say that the "observer linearization problem" (OLP) is *solvable* for the model (1) if we can find $U_0$, $\mathbf{x}_0 \in U_0$, $\Phi : U_0 \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\alpha : h(U_0) \rightarrow \mathbb{R}^2$, such that the model (1) is transformed into (5), with $(C, A)$ observable, for all $\mathbf{z} \in \Phi(U_0)$. A first necessary condition in order to be able to transform the model into observer form is its local (weak) observability, which is – in our 2-dimensional case – $\dim(\mathrm{span}\{dh\ ,\ dL_f h\}_{|\mathbf{X}}) = n = 2\ \forall \mathbf{x} \in U_0$. When the model is locally observable, we can define $\tau$ as the *unique* vector field on $U_0$ that satisfies

$$\begin{cases} L_\tau h(\mathbf{x}) = 0 \\ L_\tau L_f h(\mathbf{x}) = 1 \end{cases} \qquad \forall\ \mathbf{x} \in U_0. \tag{8}$$

If one can find a diffeomorphism $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ mapping $\mathbf{x}$ into $\mathbf{z}$ such that

$$\frac{\partial F}{\partial \mathbf{z}} = [\tau\ \ L_f\tau]\,, \tag{9}$$

then it is easy to check that $\Phi \doteq F^{-1}$ and $\alpha(\mathbf{z}) = \left[\frac{\partial \Phi}{\partial \mathbf{X}} f(\mathbf{x})\right]_{\Phi^{-1}(\mathbf{Z})} = \left[\ 0 \quad \mathbf{z}_1\ \right]$ solve the observer linearization problem (Isidori, 1989).

Therefore the solution to the OLP boils down to the solution of the partial differential equation (PDE) of eq. (9). In order to study the existence of a solution, it is not necessary to solve explicitly the PDE, for there is an equivalent condition expressed only in terms of the vector field $\tau$. In fact (Isidori, 1989) the OLP is solvable for the planar SFM problem *if and only if*

$$\begin{cases} 1)\ \dim(\mathrm{span}\{dh\ ,\ dL_f h\}_{\mathbf{X}}) = 2 \\ 2)\ \tau \text{ is such that } [L_f^i\tau\ ,\ L_f^j\tau] = 0\ \forall\ i, j\ = 0, 1 \end{cases} \tag{10}$$

where [ , ] denotes the Lie bracket of two vector fields. In the case of planar structure from motion, we have

$$\begin{bmatrix} dh \\ dL_f h \end{bmatrix} = \begin{bmatrix} \frac{1}{\mathbf{X}_2} & -\frac{\mathbf{X}_1}{\mathbf{X}_2^2} \\ -\frac{v_2 + 2\omega\mathbf{X}_1}{\mathbf{X}_2^2} & \frac{2v_2\mathbf{X}_1 + 2\omega\mathbf{X}_1^2 - v_1\mathbf{X}_2}{\mathbf{X}_2^3} \end{bmatrix}. \tag{11}$$

Since the normal rank of the rightmost matrix, which is defined for $\mathbf{x}_2 \neq 0$, is 2 for $v \neq 0$, we conclude that SFM is locally (weakly) observable anywhere away from the center of projection and the necessary observability conditions 1) are met. However, condition 2) does not hold. In fact, by solving equation (8) we get

$$\tau_1 = \frac{\mathbf{x}_1}{\mathbf{x}_2}\tau_2 \tag{12}$$

$$\tau_2 = \frac{\mathbf{x}_2^3}{v_2\mathbf{x}_1 - v_1\mathbf{x}_2} \tag{13}$$

and it is immediate to verify that $[\tau\ ,\ L_f\tau] \neq 0$, therefore we conclude that *the observer linearization problem is* **not** *solvable in the case of structure from motion.*

It is possible that a more general transformation allowing also a change of output coordinates

$$\psi : \mathbb{RP}^1 \rightarrow \mathbb{RP}^1$$

$$y \mapsto z = \psi(y) \tag{14}$$

can transform the SFM model into observer form. This problem was first studied by (Krener and Respondek, 1985). A necessary condition for the existence of the change of output coordinates $\psi$ is the so-called "degree condition" that must be satisfied by the so-called "observable forms" of the model (Krener and Respondek, 1985; Phelps, 1991). The observable forms of a locally weakly observable system can be computed by changing the state coordinates using the basis $\{L_f^i h_j\}_{i=0,\ldots,n_j-1, j=1,\ldots,p}$ where $n_j$ is the observability index of the $j^{th}$ output component. For a single output system, as in the planar SFM case, $p = 1$ and $n_1 = n$ and the observable form is simply

$$\begin{aligned} y &= C\xi \\ \dot{\xi} &= A\xi + B\beta(\xi) \end{aligned}$$

where $A$ $B$ $C$ are a Brunowski form and $\beta$ is some smooth function of the state. Associating to each state component $\xi_j$ the degree $j-1$, corresponding to the number of times one must differentiate the output to obtain $\xi_j$, and to each product $\xi_i\xi_j$ the degree $i-1+j-1$ equal to the sum of the degrees of each factor, we define $\mathcal{P}^k$ to be the polynomials of degree $k$ or less with coefficients that are $\mathcal{C}^\infty$ functions of the output $\psi$ and $\mathcal{P}_0^k$ the subset of $\mathcal{P}^k$ generated by the elements of $\mathcal{P}^{k-1}$. Then, the degree condition consists of the following (Krener and Respondek, 1985): if there exists an observer form then, in any observable form, the vectorfields $\beta_j(\xi)$ are in $\mathcal{P}^{n_j}$. An immediate consequence of the above results is that, for a simple system that has observability indices equal to 1, there must exist an observable form with $\beta \in \mathcal{P}_0^1$. In fact, since $\mathcal{P}_0^1 = \mathcal{P}^0$, there must exist an observable form with the nonlinearities $\beta$ that are functions only of the output $y$. This particular observable form is already an observer form. In other words, an observer form can be computed without change of output coordinates. In the single output case, this form must coincide, obviously, with the one computed by the technique of (Krener and Isidori, 1983) described previously. The conclusion is that – in the single output case with $n = 2$ – if the technique of Krener and Isidori fails then, even allowing a change of output coordinates, the model cannot be put in observer form. Hence, for the structure from motion problem there is no observer with state space of dimension 2 and linear and spectrally assignable error dynamics.

However, there may still exist a linear system with a higher dimensional state space, which generates the same output trajectories. This problem was studied by Levine and Marino (Levine and Marino, 1986) together with a technique for solving the observer linearization problem by "linear immersion". A necessary condition for the existence of a linear immersion is that the vector field generated by $\{L_f^i h\}_{\forall i}$ is finite-dimensional. Unfortunately, this condition is not satisfied in the case of planar structure from motion. If, for the sake of simplicity, we let $\omega = 0$, it can be seen that

$$L_f^i h = -(i-1)2\frac{v_2}{x_2}L_f^{i-1}h \quad i > 1 \tag{15}$$

Therefore, the span of $\{L_f^i h\}_{\forall i}$ is infinite-dimensional since there does not exist an integer $p$ such that a linear combination of $\{L_f^i h\}_{i=0,\dots,p}$ equals zero, even allowing the coefficients to be functions of the output.

The overall conclusion is that *the structure from motion problem*, apart from some particular trivial configurations, *is an intrinsically nonlinear observation problem and there does not exist an observer on any (arbitrarily large) embedding space that has linear and spectrally assignable error dynamics.* Clearly, in the particular and trivial case in which $\omega = 0$ and $v_2 = 0$, $L_f^i h = 0$ for $i \geq 2$, $\mathbf{x}_2 =$ constant and there exists an optimal filter with linear error dynamics; it is the Kalman filter for the model

$$\begin{cases} y = \xi_1 \\ \dot{\xi}_1 = \xi_2 \\ \dot{\xi}_2 = 0 \end{cases}$$
$$\frac{1}{x_2} \doteq \frac{\xi_2}{v_1}. \tag{16}$$

If $x_2$ is not constant, but it has slow dynamics, it may be considered as an unknown parameter in the model, and an adaptive observer scheme may be implemented in order to estimate it. This leads to the analysis of alternative nonlinear observer schemes.

## 3. ALTERNATIVE OBSERVERS

In the literature, alternative nonlinear observer schemes have been proposed; in particular, the adaptive observer proposed by (Bastin and Gevers, 1988) and further studied by (Marino, 1990) can be easily applied with some appropriate modifications to the SFM problem. In the planar structure from motion case, the derivative of the output $y$ satisfies the differential equation

$$\dot{y} = -\omega(1+y^2) + \frac{1}{\mathbf{x}_2}(v_1 - yv_2). \tag{17}$$

If $1/\mathbf{x}_2$ was an unknown constant parameter, (17) would be exactly in the form considered in (Marino, 1990). It is easy to show using the Lyapunov function

$$V = k(y-\hat{y})^2 + k(\frac{1}{\mathbf{x}_2} - \frac{1}{\hat{\mathbf{x}}_2})^2, \tag{18}$$

where $k$ is a positive constant, that the following

$$\begin{cases} \dot{\hat{y}} = -\omega(1+y^2) + \frac{1}{\hat{\mathbf{x}}_2}(v_1 - yv_2) + k(y-\hat{y}) \\ \frac{d}{dt}\left(\frac{1}{\hat{\mathbf{x}}_2}\right) = (v_1 - yv_2)(y-\hat{y}) \end{cases} \tag{19}$$

is a globally asymptotically convergent observer since $\dot{V} = -2k^2(y-\hat{y})^2$. However, $1/\mathbf{x}_2$ is not a constant parameter, as it has its own dynamics

$$\frac{d}{dt}\left(\frac{1}{\mathbf{x}_2}\right) = -\omega y\frac{1}{\mathbf{x}_2} - v_2(\frac{1}{\mathbf{x}_2})^2 \tag{20}$$

and the convergence of (19) cannot be guaranteed by the above argument anymore. Consider then the following estimator, which is obtained from the previous one by adding the dynamics of the inverse depth

$$\begin{cases} \dot{\hat{y}} = -\omega(1+y^2) + \frac{1}{\hat{\mathbf{x}}_2}(v_1 - yv_2) + k(y-\hat{y}) \\ \frac{d}{dt}\left(\frac{1}{\hat{\mathbf{x}}_2}\right) = (v_1 - yv_2)(y-\hat{y}) - \omega y\frac{1}{\hat{\mathbf{x}}_2} - v_2(\frac{1}{\hat{\mathbf{x}}_2})^2. \end{cases} \tag{21}$$

Some properties of the structure from motion problem come at hand in order to prove convergence. The output $y$ is bounded by the dimensions of the image plane $|y| \leq m$ and the inverse of the depth satisfies $0 < 1/\mathbf{x}_2 \leq 1/f = 1$ since we assumed unitary focal length $f$. Then, under the following conditions

$$\begin{cases} \text{if } |\hat{y}(t)| > m \text{ then } \hat{y}(t^+) = \hat{y}(t)\frac{m}{|\hat{y}(t)|} \\ \text{if } \frac{1}{|\hat{\mathbf{x}}_2(t)|} > 1 \text{ then } \hat{\mathbf{x}}_2(t^+) = \frac{\hat{\mathbf{x}}_2(t)}{|\hat{\mathbf{x}}_2(t)|} \end{cases} \tag{22}$$

the estimator described above can be easily shown to converge with a sufficiently high gain $k$, with the same Lyapunov function used previously. An observer in this form was already obtained by (Jankovic and Ghosh, to appear), who first came out with a provably convergent scheme for estimating structure from motion.

## 4. LOCAL LINEARIZATION-BASED STRUCTURE FROM MOTION

The model (1), which *defines* the structure from motion problem, is not only locally weakly observable, but also its linearization about the current state is observable away from the center of projection. This is a very favorable situation for using local linearization-based observers, as for example the Extended Kalman Filter (EKF) (Jazwinski, 1970) based upon the model (1). This is essentially the approach taken in (Matthies *et al.*, 1989; Oliensis and Inigo-Thomas, 1992; Soatto *et al.*, 1993), and it has proven effective in most practical situation, when the *motion of the viewer is known.* In the experimental section we will present some comparative experiments with an EKF based upon the three-dimensional version of (1), which is

$$\begin{cases} n = 3 \\ \mathbf{x} = [\mathbf{x}_1\ ,\ \mathbf{x}_2\ ,\ \mathbf{x}_3]^T \\ f(\mathbf{x}) = \hat{\Omega}\mathbf{x} + V \\ \hat{\Omega} \doteq \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \\ V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \\ \mathbf{y} = h(\mathbf{x}) = \frac{\mathbf{x}}{\mathbf{x}_3} \qquad \mathbf{x}_3 \neq 0. \end{cases} \tag{23}$$

## 5. MOTION-INDEPENDENT STRUCTURE ESTIMATION

All we have said so far can be applied only when the relative motion between the scene and the viewer is *known* and has *constant velocity.* In many practical instances, however, this is not the case. The literature proposes a variety of motion estimation schemes which do not depend upon the structure of the scene, both from two views (see (Faugeras, 1993) for a review) and recursively from an image sequence (Soatto *et al.*, 1994). However, errors in the reconstructed motion – if not treated properly – have dramatic effect in the estimates of structure using the schemes described above, as we will see in the experimental section.

In this section we present a factorization method for recursively estimating structure *independent of motion.*

Consider the derivative of the output of the basic model (1) in its three-dimensional version (23) for a number of points $i = 1 \dots N$:

$$\dot{\mathbf{y}}_i(t) = \left[ \frac{1}{\mathbf{x}_3(t)_i} \mathcal{A}(\mathbf{y}_i)\ \mathcal{B}(\mathbf{y}_i) \right] \begin{bmatrix} V(t) \\ \Omega(t) \end{bmatrix}, \tag{24}$$

where

$$\mathcal{A}(\mathbf{y}_i) \doteq \begin{bmatrix} 1 & 0 & -\frac{\mathbf{x}_1}{\mathbf{x}_3}_i \\ 0 & 1 & -\frac{\mathbf{x}_2}{\mathbf{x}_3}_i \end{bmatrix}$$

$$\mathcal{B}(\mathbf{y}_i) = \begin{bmatrix} -\frac{\mathbf{x}_1}{\mathbf{x}_3}_i \frac{\mathbf{x}_2}{\mathbf{x}_3}_i & 1 + (\frac{\mathbf{x}_1}{\mathbf{x}_3}_i)^2 & -\frac{\mathbf{x}_2}{\mathbf{x}_3}_i \\ -1 - (\frac{\mathbf{x}_2}{\mathbf{x}_3}_i)^2 & \frac{\mathbf{x}_1}{\mathbf{x}_3}_i \frac{\mathbf{x}_2}{\mathbf{x}_3}_i & \frac{\mathbf{x}_1}{\mathbf{x}_3}_i \end{bmatrix}$$

depend only upon the measured function of the state $\mathbf{y}_{1i} = \frac{\mathbf{x}_1}{\mathbf{x}_3}_i$ and $\mathbf{y}_{2i} = \frac{\mathbf{x}_2}{\mathbf{x}_3}_i$. Observing $N$ points, one we may write

$$\dot{\mathbf{y}} = \mathcal{C}(\frac{1}{\mathbf{x}_{3\ 1}}, \dots, \frac{1}{\mathbf{x}_{3\ N}}, \mathbf{y})[V\ \Omega]^T$$

where

$$\mathcal{C} \doteq \begin{bmatrix} \frac{1}{\mathbf{x}_{3\ 1}}\mathcal{A}_1 & \mathcal{B}_1 \\ \vdots & \vdots \\ \frac{1}{\mathbf{x}_{3\ N}}\mathcal{A}_N & \mathcal{B}_N \end{bmatrix}.$$

Under the usual rank conditions, we may compute the least-squares approximation of $V, \Omega$ as

$$\begin{bmatrix} \hat{V} \\ \Omega \end{bmatrix} = \mathcal{C}^\dagger \dot{\mathbf{y}}$$

where $\dagger$ indicates the pseudo-inverse. Therefore, for $N > 3$, the motion field specifies the constraint

$$\dot{\mathbf{y}} = \mathcal{C}\mathcal{C}^\dagger \dot{\mathbf{y}} \Rightarrow \mathcal{C}^\perp(\frac{1}{\mathbf{x}_{3\ 1}}, \dots, \frac{1}{\mathbf{x}_{3\ N}}, \mathbf{y})\dot{\mathbf{y}} = 0,$$

where $\mathcal{C}^\perp \doteq I - \mathcal{C}\mathcal{C}^\dagger$.

Indeed, it is immediate to see that the problem of estimating the structure *independent of the motion of the viewer* can be rephrased as the problem of *identifying the following Exterior Differential System (Bryant* et al.*, 1991), with parameters on a sphere*, embedded in $\mathbb{R}^N$:

$$\begin{cases} \mathcal{C}^\perp(Z, \mathbf{y})\dot{\mathbf{y}} = 0 & Z \in \mathbf{S}^{N-1} \\ \mathbf{y}_i = \frac{\mathbf{x}}{\mathbf{x}_3}_i + n_i & \forall i = 1 \dots N \end{cases}$$

where $Z \doteq [\frac{1}{\mathbf{x}_{3\ 1}} \cdots \frac{1}{\mathbf{x}_{3\ N}}]^T$ and $n_i$ is a white, zero-mean Gaussian noise. We have implemented an identifier for the above model using the techniques in (Soatto *et al.*, 1994), and tested the scheme on synthetic image sequences, as shown in the experimental section.

## 6. EXPERIMENTS

In order to test the effectiveness of the schemes analyzed in the previous sections, we have performed a simulation experiment by generating a number of feature points between 5 and 50 in a cubic region of space of side 1 m and placed 1.5 m ahead of the observer. The points are projected onto an ideal image plane of size $500 \times 500$ pixels, corresponding to a visual field of approximately $50^o$, and noise is added to the measurements which is white, zero-mean and Gaussian with 0.1 to 1 pixel standard deviation, according to the performance of common feature detecting/tracking techniques (Barron *et al.*, 1992). The observer then moves around the cloud of points while maintaining its distance from the centroid fixed.

In order to use the adaptive observer described in section 3 and the EKF of section 4 it is necessary to input the relative motion between the viewer and the scene. We have added noise in the motion components with standard deviation ranging from 0 to 10 % of the value of the motion components. In practical applications the motion has to be either measured from sensors (as for example when the camera is mounted on a calibrated robot arm) or estimated from visual data (Soatto *et al.*, 1994), and estimation errors in the range of 1% to 10% are usually to be expected. In figure 2 we plot the norm of the relative error in the estimated depth for 20 points and for disturbances in the motion components of 0%, 2%, 5% and 10% respectively. It can be noticed that the error for the adaptive observer (dot-dash line), which is strongly correlated to start with, increases dramatically as the disturbance in the motion components increases. The EKF (dashed line) degrades more gracefully and its estimation error is significantly less correlated.

The motion-independent structure estimator proposed in section 5, unlike the other schemes, does not need knowledge of the motion parameters, and therefore the error level is constant across trials. Note that the scheme is less precise than the EKF when motion is known with infinite precision, but it outperforms the EKF as soon as the disturbance level in the motion components increases beyond 5%.

In figure 1 we show some closeups of the estimated inverse depth for the case of the EKF and the motion-independent estimator for a disturbance level in the components of motion of 5%. The performance of the two schemes is quite similar.

## 7. CONCLUSIONS

In this paper we have recalled the "observer linearization problem" as the problem of building a nonlinear observer for a nonlinear dynamical system, having an error which evolves according to a linear and spectrally assignable dynamic model. We have applied well-known results to the SFM problem to conclude that there exists no change of state and output coordinates that linearizes the error dynamics. The model of the SFM problem may not be immersed in a larger dimensional linear filter. This implies that, in order to estimate depth, it is necessary to design an adaptive observer with nonlinear error dynamics. An instance is the scheme proposed by (Jankovic and Ghosh, to appear). It turns out that, in most practical situations, a simple EKF outperforms any provably convergent scheme presented so far in the literature.
From a geometric point of view, there is no change of coordinates which structurally modifies the observer task. However, from a *computational* (numerical) point of view, the choice of the reference frame may play a crucial role, depending on the application. In each specific case (broad field of view, small apertures etc.) the user has to evaluate what is the best reference frame in terms of conditioning with respect to error in the location of the projection of the feature points in the image plane as well as in the components of motion.

Both the adaptive observer and the EKF need knowledge of the relative motion parameters, and they degrade quite dramatically when the error in the motion parameters increase beyond 10%. Furthermore, in many situations the observer's motion is not known at all. We have proposed a novel dynamic estimator of structure from image motion which is independent of the observer's motion. It is modeled as the identification of an implicit model, realized using a generalization of the EKF to implicit measurement constraints described in (Soatto *et al.*, 1994).

## 8. REFERENCES


Barron, J., D. Fleet and S. Beauchemin (1992). Performance of optical flow techniques. RPL-TR 9107. Queen's University Kingston, Ontario. Robotics and perception laboratory. Also in Proc. CVPR 1992, pp 236-242.

Bastin, G. and M. Gevers (1988). Stable adaptive observers for nonlinear time-varying systems. *IEEE trans. Aut. Control.*

Bryant, R. L., S. S. Chern, R. B. Gardner, H. L. Goldshmidt and P. A. Griffith (1991). *Exterior Differential Systems.* Mathematical Research Institute. Springer Verlag.

Faugeras, O. D. (1993). *Three dimensional vision, a geometric viewpoint.* MIT press.

Hermann, R. and A. J. Krener (1977). Nonlinear controllability and observability. *IEEE Trans. Aut. Contr. AC-22 pp. 728-740.*

Isidori, A. (1989). *Nonlinear Control Systems.* Springer Verlag.

Jankovic, M. and B. K. Ghosh (to appear). Visually guided ranging from observations of points, lines and curves via an identifier based nonlinear observer. *System and Control Letters.*

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory.* Academic Press.

Krener, A. J. and A. Isidori (1983). Linearization by output injection and nonlinear observers. *Systems and Control Letters vol. 3.*

Krener, A. J. and W. Respondek (1985). Nonlinear observers with linearizable error dynamics. *SIAM J. Control and Optimization.*

Lee, W. and K. Nam (1991). Observer design for autonomous discrete-time nonlinear systems. *Systems and Control Letters vol. 17.*

Levine, J. and R. Marino (1986). Nonlinear system immersion, observers and finite-dimensional filters. *Systems and Control Letters.*

Marino, R. (1990). Adaptive observers for single output nonlinear systems. *IEEE Trans. Aut. Control (35) 9.*

Matthies, L., R. Szeliski and T. Kanade (1989). Kalman filter-based algorithms for estimating depth from image sequences.. *Int. J. of computer vision.*

Nijmeijer, H. (1982). Observability of autonomous discrete time nonlinear systems. *Int. J. Control vol. 36 (5).*

Oliensis, J. and J. Inigo-Thomas (1992). Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop.*

Phelps, A. (1991). On constructing nonlinear observers. *SIAM J. of Control and Optim. (29) 3.*

Schaft, A. J. Van Der (1982). Observability and controllability for smooth nonlinear systems. *SIAM J. Control and Optim. vol. 20 (3).*

Shekhar, C. and R. Chellappa (1992). Passive ranging using a moving camera.. *J. of Robotics S. vol. 9.*

Soatto, S. (1994). Observability/identifiability of rigid motion under perspective projection. In: *33 IEEE Conf. on Decision and Control.* pp. 3235–3240. Extended version submitted to the IFAC Journal "Automatica".

Soatto, S., P. Perona, R. Frezza and G. Picci (1993). Recursive motion and structure estimation with complete error characterization. In: *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR.* pp. 428–433.

Soatto, S., R. Frezza and P. Perona (1994). Motion estimation on the essential manifold. In: *Proc. $3^{rd}$ Europ. Conf. Comput. Vision, J.-O. Eklundh (Ed.), LNCS-Series Vol. 800-801, Springer-Verlag.* pp. II–61–72.

Zhang, Z. and O. Faugeras (1992). *3D dynamic scene analysis.* Vol. 27 of *Information Sciences.* Springer-Verlag.

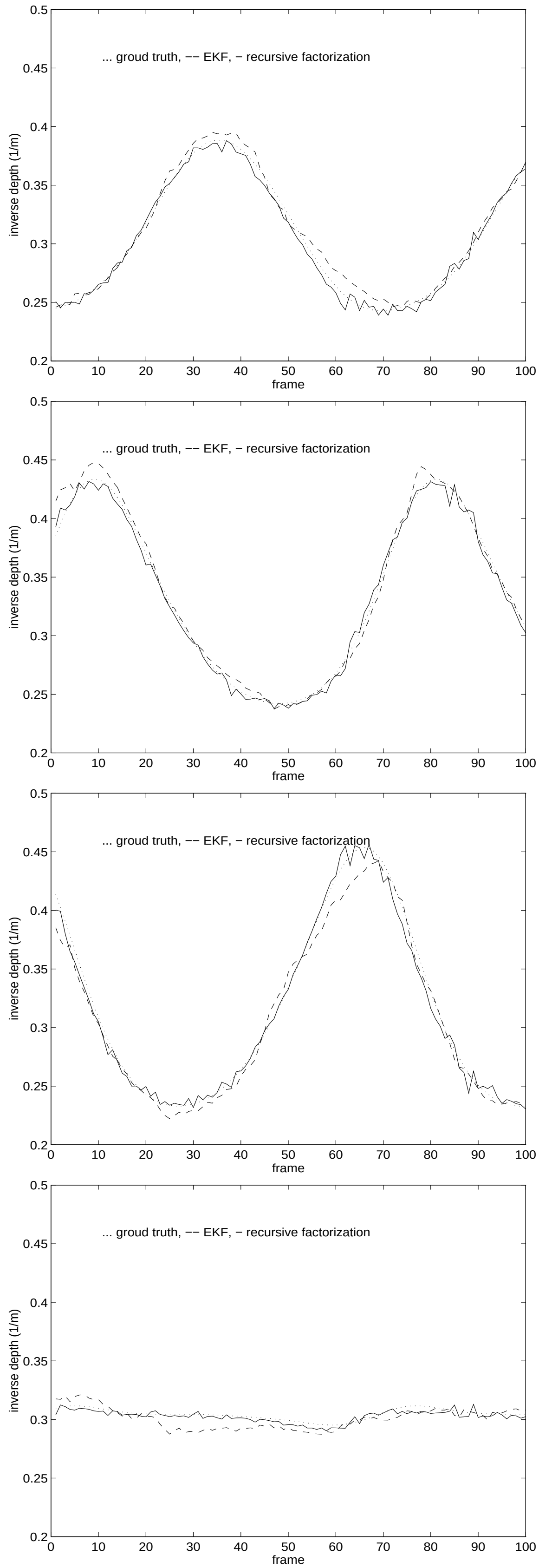


Fig. 1. *Close-ups of the estimates for the case of the EKF and the recursive motion-independent factorization method. Noise in the motion parameters for the EKF was 5%.*

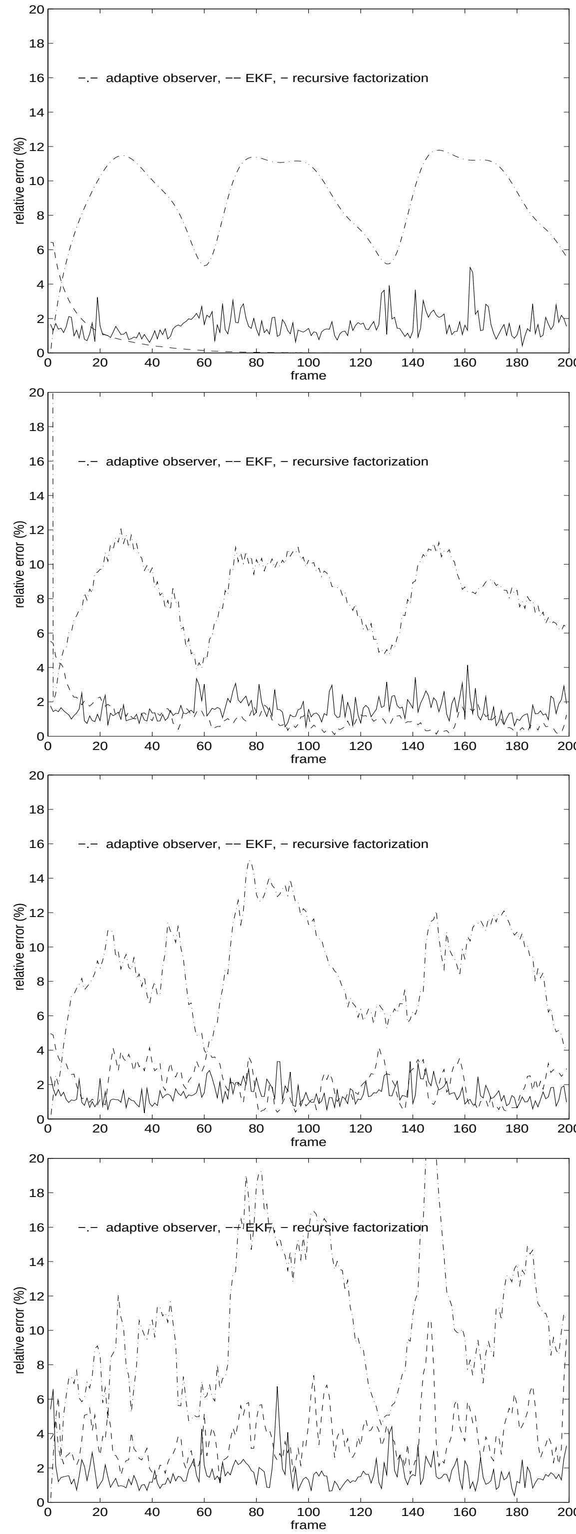


Fig. 2. *Relative error norm for the three schemes. For both the adaptive observer (dot-dash line) and the EKF (dashed line), motion needs to be known. The four plots show the relative error in presence of uncertainty in the motion parameters of (top to bottom) 0%, 2%, 5% and 10%. The novel factorization scheme presented is instead independent of the motion of the viewer, and therefore is not affected by the uncertainty in the motion parameters (solid line).*